Semiparametric Single Index Versus Fixed Link Function Modelling

W. Härdle V. Spokoiny S. Sperlich

February 21, 1996

Abstract

Discrete choice models are frequently used in statistical and econometric practice. Standard models such as logit models are based on exact knowledge of the form of the link and linear index function. Semiparametric models avoid possible misspecification but often introduce a computational burden. It is therefore interesting to decide between approaches. Here we propose a test of semiparametric versus parametric single index modelling. Our procedure allows that the (linear) index of the semiparametric alternative is different from that of the parametric hypothesis. The test is proved to be rate-optimal in the sense that it provides the (rate) minimal distance between hypothesis and alternative for a given power function.

1 Introduction

Discrete choice models are frequently used in statistical and econometric applications. Among them binary response models, such as Probit or Logit regression, dominate the applied literature. A basic hypothesis made there is that the link and the index function have a known form, see McCullagh and Nelder (1989). The fixed form of the link function e.g. the logistic cdf is rarely justified by the context of the observed data but is often motivated by numerical convenience and by reference to "standard practice", say "accessible canned software".

Recent theoretical and practical studies have questioned this somewhat rigid approach and have proposed a more flexible semiparametric approach. Green and Silverman (1994) use the theory of penalizied likelihood to model nonparametric link functions with splines. Horowitz (1993) gives an excellent survey on single index methods and stresses economic applications. Staniswalis and Severini (1994) use kernel methods and keep a fixed link function but allow the index to be of partial linear form. Partial linear models are semiparametric models with a parametric linear and a nonparametric index and have been studied by Rice (1986), Speckman (1988) and Engle, Granger, Rice and Weiss (1986).

These models enhance the class of Generalized Linear Models (McCullagh and Nelder, 1989) in several ways. Here we concentrate on one generalization, the single index models with link function of unknown nonparametric form but (linear) index function. The advantage of this approach is that still an interpretable linear single index, a weighted sum of the predictor variables, is produced. The link function plays in theoretical justifications of single index models via stochastic utility functions an important role (Maddala, 1983): it is the cdf of the errors in a latent variable model. Our approach enables us to interpret the results still in terms of a stochastic utility model but enhances it by allowing for an unknown cdf of the errors.

Despite the gained flexibility in semiparametric regression modelling there is still an important gap between theory and practice, namely a device for testing between a parametric and semiparametric alternative. A first paper in bridging this gap is Horowitz and Härdle (1994). They considered for response Y and predictor X the parametric null hypothesis

$$H_0 : Y = F(X^\top \theta_0) + \varepsilon \tag{1}$$

where $x^\top \theta$ denotes the index and F is the fixed and known link function. The semiparametric alternative considered there is that the regression function has the form $f(x^\top \theta_0)$ with a nonparametric link function f and the same index $x^\top \theta_0$ as under H_0. The main drawback of that paper is that the index is supposed to be the same under the null and the alternative.

The goal of the present paper is to construct a test which has power for as large class of alternatives. We move to a full semiparametric alternative by considering alternatives of single index type

$$H_1 : Y = f(X^\top \beta) + \varepsilon \tag{2}$$

with β possibly different from θ_0. The situation of our test is illustrated in the following figures 1 and 2.

The data is a crosssection of 462 records on apprenticeship of the German Social Economic Panel from 1984 to 1992. The dependent variable is an indicator of unemployement, ($Y = 1$ =yes). Explanatory variables are X_1 gross monthly earnings as an apprentice, X_2 percentage of people apprenticed in a certain occupation, divided by the people employed in this occupation in the entire economy and X_3 unemployment rate in the state the respondent lived in during the year the apprenticeship was completed. The aim of the test is to decide between the logit model and the semiparametric model with unknown link function and possibly different index. In Härdle, Klinke and Turlach (1995) this hypothesis is tested with the Horowitz Härdle test by Proenca and Werwatz who also prepared the dataset. They give a delicious description of the test procedure but it does not reject.

We measure the quality of a test by the value of minimal distance between the regression function under the null and under the alternative which is sufficient to provide the desir-

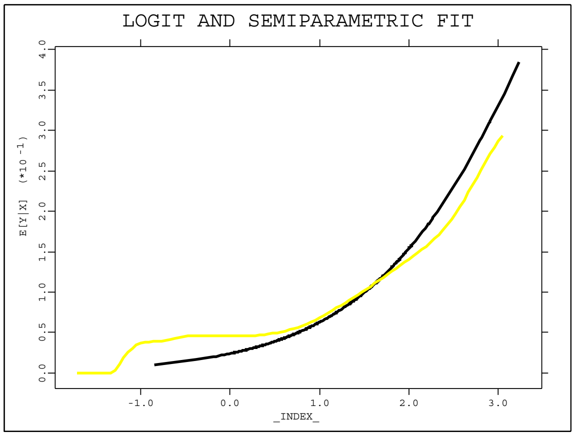

Figure 1: Parametric fitting

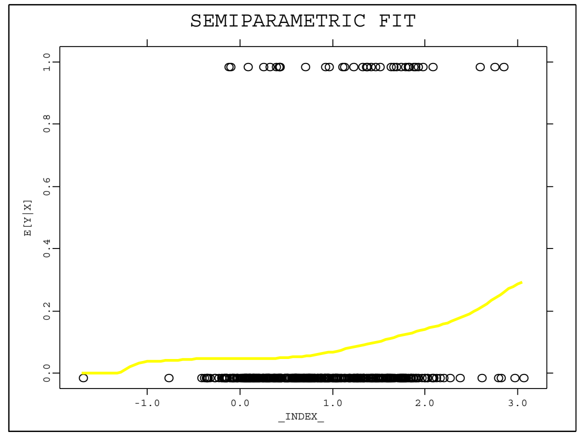

Figure 2: Semiparametric fitting

able power of testing. The test proposed below is shown to be rate-optimal in this sense. The paper is organized as follows. The next section contains the main results then we present the test procedure. In Section 5 we present some simulation study. The proof of main results are given in Section 3 (Theorem 2.2) and in the Appendix (Theorem 2.1).

2 Main Results

We start with a brief historical background of the nonparametric hypothesis testing problem. The problem for the case of a simple hypothesis and univariate nonparametric alternative was considered by Ibragimov and Khasminskii (1977) and Ingster (1982). It was shown that the minimax rate for the distance between the null and the alternative set is of the order $n^{-2s/(4s+1)}$ where s is a measure of smoothness. Note that this rate differs from that of an estimation problem where we have $n^{-s/(2s+1)}$. In the multivariate case the corresponding rate changes to $n^{-2s/(4s+d)}$, as Ingster (1993) has shown. The problem of testing a parametric hypothesis versus a nonparametric alternative was discussed also in Härdle and Mammen (1993). Their results allow to extract the above minimax rate.

The results of Friedman and Stuetzle (1981), Huber (1985), Hall (1989) and Golubev (1992) show that estimation of the function f under (2) can be made with the rate corresponding to the univariate case. Below we will see though that for the problem of hypothesis testing the situation is slightly different. The rate for this additive alternative of single index type differs from that of a univariate alternative ($d=1$) by an extra log-factor. Nevertheless, we have almost a univariate rate and we can therefore still expect efficiency of the test for practical applications.

We will come back to the introductory example in section 5. Suppose we are given independent observations (X_i, Y_i), $X_i \in I\!R^d$, $Y_i \in I\!R^1$, $i = 1, \ldots, n$, that follow the regression

$$Y_i = F(X_i) + \varepsilon_i, \qquad i = 1, \ldots, n. \tag{3}$$

Here $\varepsilon_i = Y_i - F(X_i)$ are mean zero error variables,

$$\mathbf{E}\varepsilon_i = 0, \qquad i = 1, \ldots, n,$$

with conditional variance

$$\sigma_i^2 = \mathbf{E}\left[\varepsilon_i^2 \mid X_i\right], \qquad i = 1, \ldots, n. \tag{4}$$

Example **2.1** As a first example take the above single index binary choice model. The observed response variables Y_i take two values $0, 1$ and

$$\begin{aligned}\mathbf{P}\left(Y_i = 1 \mid X_i\right) &= F(X_i),\\ \mathbf{P}\left(Y_i = 0 \mid X_i\right) &= 1 - F(X_i).\end{aligned}$$

In this case $\sigma_i^2 = F(X_i)\{1 - F(X_i)\}$.

Example **2.2** A second example is a nonlinear regression model with unknown transformation. An excellent introduction into nonlinear regression can be found in Huet, Jolivet and Messeau (1993). The model takes the same form as (1) but the response Y is not necessarily binary and the variance σ_i^2 may be an unknown function of the $F(X_i)$'s. Carroll and Ruppert (1988) use this kind of error structure to model fan shaped residual structure.

We wish to test the hypothesis H_0 that the regression function $F(x)$ belongs to a prescribed parametric family $(F_\theta(x),\, \theta \in \Theta)$, where Θ is a subset in a finite-dimensional space $I\!R^m$. This hypothesis is tested versus the semiparametric alternative H_1 that the regression function $F(\cdot)$ is of the form

$$F(x) = f(x^\top \beta) \tag{5}$$

where β is a vector in $I\!R^d$ with $|\beta| = 1$, and $f(\cdot)$ is a univariate function.

Example **2.3** Let the parametric family $(F_\theta(x), \theta \in \Theta)$ be of the form

$$F_\theta(x) = \frac{1}{1 + \exp(-x^\top \theta)} \tag{6}$$

and let otherwise (X, Y) have stochastic structure as in Example 2.1. This form of parametrization leads to a binary choice logit regression model. Probit or complementary log-log models have a different parametrization but still have this single index form.

Let $\mathcal{F}_0$ be the set of functions $(F_\theta(x),\, \theta \in \Theta)$ and let $\mathcal{F}_1$ be a set of alternatives of the form (5). We measure the power of a test φ_n by its power function on the sets $\mathcal{F}_0$ and $\mathcal{F}_1$: if $\varphi_n = 0$ then we accept the hypothesis H_0 and if $\varphi_n = 1$, then we accept H_1. The corresponding first and second type error probabilities are defined as usual:

$$\alpha_0(\varphi_n) = \sup_{F \in \mathcal{F}_0} \mathbf{P}_F\left(\varphi_n = 1\right),$$

$$\alpha_1(\varphi_n) = \sup_{F \in \mathcal{F}_1} \mathbf{P}_F\left(\varphi_n = 0\right).$$

Here $\mathbf{P}_F$ means the distributions of observations (X_i, Y_i) given the regression function $F(\cdot)$. When there is no risk of confusion we write $\mathbf{P}$ instead of $\mathbf{P}_F$. Our goal is to construct a test φ_n that has power over a wide class of alternatives. The assumptions needed are made precise below. We start with assumptions on the error distribution.

$(E1)$ The errors ε_i are bounded by a universal constant C_ε

$$\varepsilon_i \le C_\varepsilon\,, \qquad i = 1, \ldots, n.$$

$(E2)$ The conditional distributions of errors ε_i given X_i depend only on values of the regression function $F(X_i)$,

$$\mathcal{L}\left(\varepsilon_i \mid X_i\right) = \mathcal{L}\left(\varepsilon_i \mid F(X_i)\right) = P_{F(X_i)}$$

where (P_z) is a prescribed distribution family of one-dimensional parameter z;

$(E3)$ The variance function $\sigma^2(z) = \mathbf{E}\left[\varepsilon_i^2 \mid F(X_i) = z\right]$ and the fourth central moment function $\kappa^4(z) = \mathbf{E}\left[\left(\varepsilon_i^2 - \mathbf{E}\varepsilon_i^2\right)^2 \mid F(X_i) = z\right]$ are separated away from zero and infinity i.e.

$$0 < \sigma_* \le \sigma(z) \le \sigma^* < \infty$$

$$0 < \kappa_* \le \kappa(z) \le \kappa^* < \infty$$

with some prescribed $\sigma_*, \sigma^*, \kappa_*, \kappa^*$, and this function is uniformly continuous: for some positive constants C_σ and C_κ one has

$$\left|\sigma(z) - \sigma(z')\right| \le C_\sigma \left|z - z'\right|,$$

$$\left|\kappa(z) - \kappa(z')\right| \le C_\kappa \left|z - z'\right|.$$

Note that $(E1)$ is obviously fulfilled for the single index model in Examples 2.1 and 2.3. In the more general situation of Example 2.2 this assumption can be weakened to the existence of exponential moments for ε_i.

The assumption $(E3)$ restricts the set of X-observations to a bounded set. It is made more precise in the following assumption on the design X.

(D) The predictor variables X have a design density $\pi(x)$ which is supported on the compact convex set $\mathcal{X}$ in $\mathbb{R}^d$ and is separated from zero and infinity on $\mathcal{X}$;

Assumption (D) is quite common in nonparametric regression analysis. It is apparently fulfilled for the above example on apprenticeship and youth unemployment. We now specify the hypothesis and alternative.

$(H0)$ The parameter set Θ is a compact subset in $\mathbb{R}^m$.

For some universal constant C_Θ the following holds

$$|F_\theta(x) - F_{\theta'}(x)| \le C_\Theta \, |\theta - \theta'| \, , \qquad \forall x \in \mathcal{X},\ \theta, \theta' \in \Theta;$$

All functions $F_\theta(\cdot)$ belong to the Hölder class $\Sigma_d(s, L)$ of functions in $\mathbb{R}^d$.

$(H1)$ The univariate link function $f(\cdot)$ from (5) belongs to the Hölder class $\Sigma(s, L)$. The function $F(x) = f(x^\top \beta)$ is separated away from the parametric family $\mathcal{F}_0$ i.e.

(7) $$\inf_{\theta \in \Theta} \|F - F_\theta\| \ge c_n$$

with a given $c_n > 0$. Here $\|F - F_\theta\| = \int |F(x) - F_\theta(x)|^2 \, \pi(x) \, dx$.

For the definition of a Hölder smoothness class in the context of statistical nonparametric problems we refer e.g. to Ibragimov and Khasminskii (1981). Assumption $(H0)$ is certainly fulfilled for Example 2.3 but also in Probit and other generalized linear regression models such as the log linear models.

The main results are given below. We compute first the optimal rate of convergence of the distance c_n distinguishing the null from the alternative. The second theorem states the existence of an optimal test. The test will be given more explicitly in the next section where we also apply it to the above concrete examples. Theorem 2.2 is proved in Section 4 and the proof of Theorem 2.1 is given in the appendix.

Theorem 2.1 *Let $c_n = \left(a \frac{\sqrt{\ln n}}{n}\right)^{\frac{2s}{4s+1}}$. If a is small enough then for any sequence of tests φ_n one has*

$$\liminf_{n \to \infty} \alpha_0(\varphi_n) + \alpha_1(\varphi_n) \ge 1.$$

Theorem 2.2 *For any constant a^* large enough there is a sequence of tests φ_n^* which distinguish consistently the hypothesis H_0 versus alternative $H_1 = H_1(c_n^*)$ with $c_n^* = \left(a^* \frac{\sqrt{\ln n}}{n}\right)^{\frac{2s}{4s+1}}$ i.e.*

$$\lim_{n \to \infty} \alpha_0(\varphi_n^*) = 0$$

and

$$\lim_{n \to \infty} \alpha_1(\varphi_n^*) = 0.$$

3 The test procedure

Before we describe the test procedure let us introduce some notation. Given functions $F(x)$ and $G(x)$ we denote by

(8) $$\langle F, G \rangle = \frac{1}{n} \sum_{i=1}^{n} F(X_i) \, G(X_i).$$

the scalar product of the functions F and G. We write also $\langle F \rangle$ instead of $\langle F, F \rangle$ and identify the sequences $(Y_i), (\varepsilon_i)$ with the functions $Y(X_i)$ and $\varepsilon(X_i)$. We construct the tests φ_n^* from Theorem 2.2 in several steps.

First we shall do a preliminary pilot estimation $\tilde{F}_0$ under the null. Second we estimate the d-dimensional nonparametric regression $\tilde{F}_1$ necessary to construct estimators of expected value and the variance of the proposed test statistics. In the third step we estimate for each feasible value of β the corresponding link function f under $(H1)$ as in (2). Finally we compute the test statistic based on comparison of residuals under H_0 and H_1.

3.1 Parametric pilot estimation

Let Θ_n be a grid in the parametric set Θ with the step $\frac{\ln n}{\sqrt{n}}$. Put

$$\tilde{\theta}_n = \operatorname*{arginf}_{\theta \in \Theta_n} \langle Y - F_\theta \rangle = \operatorname*{arginf}_{\theta \in \Theta_n} \frac{1}{n} \sum_{i=1}^{n} |Y_i - F_\theta(X_i)|^2 . \tag{9}$$

Denote also

$$\tilde{F}_0(\cdot) = F_{\tilde{\theta}_n}(\cdot). \tag{10}$$

Note that $\tilde{\theta}_n$ is not necessarily an efficient estimator under the null since we do not correct for the variance function.

3.2 Nonparametric pilot estimation

For the nonparametric estimation of the expected value and the variance of the test statistic we shall use the standard kernel technique, see e.g. Härdle (1990) or Müller (1987). More precisely we use a one dimensional kernel satisfying the conditions

$(K1)$ $K(\cdot)$ is compactly supported;

$(K2)$ $K(\cdot)$ is symmetric;

$(K3)$ $K(\cdot)$ has s continuous derivatives;

$(K4)$ $\int K(t)\, dt = 1$;

$(K5)$ $\int K(t)\, t^k dt = 0, \quad k = 1, \ldots, s-1.$

Recall from $(H0)$ and $(H1)$ that s denotes the degree of smoothness of the regression function. Note also that $(K5)$ ensures that K is orthogonal to polynomials of order

1 to $s-1$. For a list of kernels satisfying $(K1)-(K5)$ we refer to Müller (1987). A d-dimensional product kernel K_1 is defined as

$$K_1(u_1,\dots,u_d)=\prod_{j=1}^{d}K(u_j). \tag{11}$$

Take now

$$h_1=n^{-\frac{1}{2s+d}}, \tag{12}$$

the optimal smoothing bandwidth in d-dimensions, and put

$$\tilde{F}_1(x)=\frac{\sum_{i=1}^{n}Y_iK_1\left(\frac{x-X_i}{h_1}\right)}{\sum_{i=1}^{n}K_1\left(\frac{x-X_i}{h_1}\right)}. \tag{13}$$

The nonparametric kernel smoother $\tilde{F}_1$ is the well known multidimensional Nadaraya-Watson kernel estimator.

3.3 Estimation under H_1

Set

$$h=\left(\frac{\sqrt{\ln n}}{n}\right)^{\frac{2}{4s+1}}. \tag{14}$$

We will use this bandwidth for estimation in the semiparametric model. Note that in (12) for the nonparametric estimation problem another rate, namely $n^{-1/(2s+d)}$ was used. Here we have almost this bandwidth except for the extra log-term.

Let S_d be the unit sphere in $I\!R^d$. Denote by $S_{n,d}$ a discrete grid in S_d with the step $b_n=h^{2s+2}$. Let N be the cardinality of $S_{n,d}$

$$N=\#S_{n,d}. \tag{15}$$

For each $\beta\in S_{n,d}$ define

$$K_{h,\beta}(x)=K\left(\frac{x^\top\beta}{h}\right),\quad x\in I\!R^d, \tag{16}$$

and introduce the smoothing operator $\mathcal{K}_\beta$ with

$$\mathcal{K}_\beta Y(X_i)=\Pi_\beta(X_i)\sum_{j\neq i}Y_jK_{h,\beta}(X_i-X_j) \tag{17}$$

where

$$\Pi_\beta(X_i)=\left(\sum_{j\neq i}K_{h,\beta}(X_i-X_j)\right)^{-1}. \tag{18}$$

Similarly we define $\mathcal{K}_\beta\varepsilon$ and $\mathcal{K}_\beta F$. Note that given β the values $\mathcal{K}_\beta Y$ estimate f in (2).

3.4 The test statistic

Now for each β we calculate a statistic T_β as follows:

$$T_\beta = \frac{n\sqrt{h}}{\tilde{V}_\beta}\left[2\left\langle Y - \tilde{F}_0, \mathcal{K}_\beta Y - \tilde{F}_0\right\rangle - \left\langle \mathcal{K}_\beta Y - \tilde{F}_0\right\rangle + \tilde{E}_\beta\right]. \tag{19}$$

Here $\langle\cdot\rangle$ is defined by (8), h by (14), $\tilde{F}_0$ by (10). We use the following notation

$$\tilde{E}_\beta = \frac{1}{n}\sum_i\sum_{j\neq i}\tilde{\sigma}_j^2\,\Pi_\beta^2(X_i)\,K_{h,\beta}^2(X_i - X_j) \tag{20}$$

where $\Pi_\beta(X_i)$ is from (18),

$$\tilde{\sigma}_j^2 = \sigma^2\left(\tilde{F}_1(X_j)\right), \quad j = 1,\ldots,n, \tag{21}$$

the function $\sigma^2(\cdot)$ being defined in the model assumptions and $\tilde{F}_1(x)$ being the nonparametric pilot estimator. Finally,

$$\begin{aligned}\tilde{V}_\beta^2 &= h\sum_i\sum_{j\neq i}\tilde{\sigma}_i^2\,\tilde{\sigma}_j^2\,\Pi_\beta^2(X_i)\left|2K_{h,\beta}(X_i - X_j) - K_{h,\beta}^{(2)}(X_i, X_j)\right|^2 + \\ &\quad + h\sum_i\tilde{\kappa}_i^4\left|\sum_{j\neq i}\Pi_\beta^2(X_j)K_{h,\beta}^2(X_i - X_j)\right|^2\end{aligned}$$

with $\tilde{\kappa}_i = \kappa\left(\tilde{F}_1(X_i)\right), \quad i = 1,\ldots,n, \qquad \kappa(\cdot)$ being from $(E3)$ and

$$K_{h,\beta}^{(2)}(X_i, X_j) = \frac{1}{\Pi_\beta(X_i)}\sum_{k\neq i,j}\Pi_\beta^2(X_k)\,K_{h,\beta}(X_k - X_i)\,K_{h,\beta}(X_k - X_j). \tag{22}$$

Put now

$$T_n^* = \sup_{\beta\in S_{n,d}} T_\beta \tag{23}$$

and

$$\varphi_n^* = \mathbf{1}\left(T_n^* > \sqrt{(2+\delta)\log N}\right). \tag{24}$$

Here $\mathbf{1}(\cdot)$ is the indicator function of the corresponding event, δ is an arbitrary small positive number and N is the cardinality of $S_{n,d}$, see (15).

4 Proof of Theorem 2

We start with the decomposition of the test statistics T_β. Denote by $B_\beta(x)$ the bias function for the smoothing operator $\mathcal{K}_\beta$ from (17):

$$B_\beta(X_i) = \mathcal{K}_\beta F(X_i) - F(X_i), \quad i = 1,\ldots,n. \tag{25}$$

Fix some $\beta \in S_{n,d}$ and $F \in \mathcal{F}_0 \cup \mathcal{F}_1$.

Lemma 4.1

$$
\begin{aligned}
T_\beta &= \frac{n\sqrt{h}}{\tilde{V}_\beta}\left[\left\langle F-\tilde{F}_0\right\rangle - \langle B_\beta\rangle + \right.\\
&\quad + 2\langle \mathcal{K}_\beta\varepsilon,\varepsilon\rangle - \langle \mathcal{K}_\beta\varepsilon\rangle + \tilde{E}_\beta + \\
&\quad \left. + 2\left\langle F-\tilde{F}_0,\varepsilon\right\rangle + 2\langle B_\beta,\varepsilon\rangle - 2\langle B_\beta,\mathcal{K}_\beta\varepsilon\rangle\right]. \qquad (26)
\end{aligned}
$$

Proof. By definition $Y = F + \varepsilon$ and therefore

$$\mathcal{K}_\beta Y = \mathcal{K}_\beta F + \mathcal{K}_\beta\varepsilon = F + B_\beta + \mathcal{K}_\beta\varepsilon.$$

Now

$$
\begin{aligned}
2\left\langle Y-\tilde{F}_0, \mathcal{K}_\beta Y - \tilde{F}_0\right\rangle &= 2\left\langle F-\tilde{F}_0+\varepsilon, F-\tilde{F}_0+B_\beta+\mathcal{K}_\beta\varepsilon\right\rangle = \\
&= 2\left\langle F-\tilde{F}_0\right\rangle + 2\left\langle F-\tilde{F}_0, B_\beta\right\rangle + 2\left\langle F-\tilde{F}_0,\mathcal{K}_\beta\varepsilon\right\rangle + \\
&\quad + 2\left\langle \varepsilon, F-\tilde{F}_0\right\rangle + 2\langle\varepsilon, B_\beta\rangle + 2\langle\varepsilon,\mathcal{K}_\beta\varepsilon\rangle
\end{aligned}
$$

and

$$
\begin{aligned}
\left\langle \mathcal{K}_\beta Y - \tilde{F}_0\right\rangle &= \left\langle F-\tilde{F}_0+B_\beta+\mathcal{K}_\beta\varepsilon\right\rangle = \\
&= \left\langle F-\tilde{F}_0\right\rangle + \langle B_\beta\rangle + \langle\mathcal{K}_\beta\varepsilon\rangle + \\
&\quad + 2\left\langle F-\tilde{F}_0, B_\beta\right\rangle + 2\left\langle F-\tilde{F}_0,\mathcal{K}_\beta\varepsilon\right\rangle + 2\langle B_\beta,\mathcal{K}_\beta\varepsilon\rangle.
\end{aligned}
$$

Substituting this in the definition of T_β we obtain the assertion of the lemma.

The next step is to show that the expansion (26) for the statistic T_β can be simplified by discarding lower order terms. Indeed we shall see below that the last three terms are relatively small and can be omitted. The terms $\tilde{E}_\beta$ and $\tilde{V}_\beta$ can be substituted by similar expressions E_β and V_β which use "true" values σ_i and κ_i instead of estimated values $\tilde{\sigma}_i$ and $\tilde{\kappa}_i$ and finally, the parametric estimator $\tilde{\theta}_n$ can be replaced by θ_n defined by

$$\theta_n = \underset{\theta\in\Theta_n}{\operatorname{arginf}} \langle F - F_\theta\rangle \qquad (27)$$

where F is a "true" regression function from (3). Suppose that all these replacements can be done. Define now

$$
\begin{aligned}
T'_\beta &= \frac{n\sqrt{h}}{V_\beta}\left[\langle F-F_{\theta_n}\rangle - \langle B_\beta\rangle + \right.\\
&\quad \left. +\ 2\langle\mathcal{K}_\beta\varepsilon,\varepsilon\rangle - \langle\mathcal{K}_\beta\varepsilon\rangle + E_\beta\right]
\end{aligned}
$$

with

$$
\begin{aligned}
E_\beta &= \frac{1}{n}\sum_i\sum_{j\neq i}\sigma_j^2\,\Pi_\beta^2(X_i)\,K_{h,\beta}^2(X_i-X_j),\\
V_\beta^2 &= h\sum_i\sum_{j\neq i}\sigma_i^2\,\sigma_j^2\,\Pi_\beta^2(X_i)\left|2K_{h,\beta}(X_i-X_j) - K_{h,\beta}^{(2)}(X_i,X_j)\right|^2 + \\
&\quad + h\sum_i\kappa_i^4\left|\sum_{j\neq i}\Pi_\beta^2(X_j)K_{h,\beta}^2(X_i-X_j)\right|^2.
\end{aligned}
$$

Below we show that the tests φ_n^{**} based on the statistics T_n^{**} with

$$T_n^{**} = \sup_{\beta \in S_{n,d}} T'_\beta \tag{28}$$

have the same asymptotic behavior as φ_n^*. For the moment we only consider the tests φ_n^{**}. Note that they are not tests in the usual sense since they use the non-observable values $E_\beta, V_\beta, \theta_n$. Central to our proof is the analysis of the asymptotic behavior of the random variables

$$\xi_\beta = n\sqrt{h}\ \left[2\left\langle \mathcal{K}_\beta \varepsilon, \varepsilon \right\rangle - \left\langle \mathcal{K}_\beta \varepsilon \right\rangle + E_\beta\right]. \tag{29}$$

Lemma 4.2 *The following assertions hold*

$$\begin{aligned} \mathbf{E}\, \xi_\beta &= 0, & (30)\\ \mathbf{E}\, \xi_\beta^2 &= V_\beta^2\,, & (31) \end{aligned}$$

and uniformly in $F \in \mathcal{F}_0 \cup \mathcal{F}_1$, $\beta \in S_{n,d}$ *and* $t \in [-\ln n, \ln n]$

$$\frac{\mathbf{P}\left(\frac{\xi_\beta}{V_\beta} > t\right)}{1 - \Phi(t)} \to 1, \quad n \to \infty, \tag{32}$$

$\Phi(\cdot)$ *being the standard normal distribution.*

Proof. The first two statements are derived by direct calculation. In fact, by definition and (22)

$$\begin{aligned} \xi_\beta \;&=\; 2\sqrt{h}\sum_i \varepsilon_i\, \Pi_\beta(X_i) \sum_{j\neq i} \varepsilon_j\ K_{h,\beta}(X_i - X_j) - \\ &\quad -\sqrt{h}\sum_i\ \Pi_\beta^2(X_i)\left|\sum_{j\neq i} \varepsilon_j\ K_{h,\beta}(X_i - X_j)\right|^2 + \\ &\quad +\sqrt{h}\sum_i \sum_{j\neq i} \sigma_j^2\, \Pi_\beta^2(X_i)\, K_{h,\beta}^2(X_i - X_j) = \\ &=\; \sqrt{h}\sum_i \sum_{j\neq i} \varepsilon_i\, \varepsilon_j\, \Pi_\beta(X_i)\left[2K_{h,\beta}(X_i - X_j) - K_{h,\beta}^{(2)}(X_i, X_j)\right] + \\ &\quad +\sqrt{h}\sum_i \sum_{j\neq i} (\sigma_j^2 - \varepsilon_j^2)\, \Pi_\beta^2(X_i)\, K_{h,\beta}^2(X_i - X_j). \end{aligned}$$

Since the errors ε_i are independent and $\mathbf{E}\, \varepsilon_i = 0$, $\mathbf{E}\, \varepsilon_i^2 = \sigma_i^2$, we immediately obtain (30) and (31). The last statement (32) is a particular case of the general central limit theorem for quadratic forms of independent random variables and can be obtained in a standard way by calculation of the corresponding cumulants. We omit the details, see e.g. Härdle and Mammen (1993).

The assertion (32) of Lemma 4.2 straightforwardly implies the following corollary.

Lemma 4.3 *Uniformly in $F \in \mathcal{F}_0 \cup \mathcal{F}_1$ one has*

$$\mathbf{P}\left(\sup_{\beta \in S_{n,d}} \frac{\xi_\beta}{V_\beta} > \sqrt{(2+\delta)\ln N}\right) \to 0, \quad n \to \infty. \tag{33}$$

Proof. For any t one gets

$$\mathbf{P}\left(\sup_{\beta \in S_{n,d}} \frac{\xi_\beta}{V_\beta} > t\right) \le \sum_{\beta \in S_{n,d}} \mathbf{P}\left(\frac{\xi_\beta}{V_\beta} > t\right) \le N \sup_{\beta \in S_{n,d}} \mathbf{P}\left(\frac{\xi_\beta}{V_\beta} > t\right). \tag{34}$$

But through (32) for n large enough

$$\mathbf{P}\left(\frac{\xi_\beta}{V_\beta} > \sqrt{(2+\delta)\ln N}\right) \le 2\left(1 - \Phi\left(\sqrt{(2+\delta)\ln N}\right)\right) \le$$
$$\le \exp\left\{-\frac{1}{2}\left|\sqrt{(2+\delta)\ln N}\right|^2\right\} = N^{-1-\delta/2}$$

that implies (33) through (34).

Now we come to the calculation of the error probabilities for the tests φ_n^{**} based on T_n^{**}. Under the hypothesis H_0 one has $F = F_\theta, \theta \in \Theta$. This does not automatically yield $\langle F - F_{\theta_n}\rangle = 0$ since $\theta_n \in \Theta_n$, see (27), and θ can be outside Θ_n. But the assumptions $(H0)$ on the parametric family guarantee that this value is small enough.

Lemma 4.4 *Let $F = F_\theta, \theta \in \Theta$. Then*

$$\langle F_\theta - F_{\theta_n}\rangle \le C_\theta^2 \frac{\ln^2 n}{n}.$$

Proof. Let

$$\theta_n' = \operatorname*{arginf}_{\theta' \in \Theta_n} |\theta - \theta'|.$$

The definition of the grid Θ_n provides $|\theta - \theta_n'|^2 \le \frac{\ln^2 n}{n}$. Now from the definition of θ_n and the assumptions $(H0)$ on the parametric family we obtain

$$\langle F_\theta - F_{\theta_n}\rangle \le \left\langle F_\theta - F_{\theta_n'}\right\rangle = \frac{1}{n}\sum_i \left|F_\theta(X_i) - F_{\theta_n'}(X_i)\right|^2 \le C_\Theta^2 |\theta - \theta'|^2 \le C_\Theta^2 \frac{\ln^2 n}{n}.$$

Using this result we have for $F = F_\theta$ by Lemma 4.3

$$\mathbf{P}\left(T_n^{**} > \sqrt{(2+\delta)\ln N}\right) \le$$
$$\le \mathbf{P}\left(\sup_{\beta \in S_{n,d}} \frac{\xi_\beta}{V_\beta} > \sqrt{(2+\delta)\ln N} - C_\theta^2 \frac{\ln^2 n}{n} n\sqrt{h}\right) \to 0, \quad n \to \infty,$$

i.e.

$$\alpha_0(\varphi_n^{**}) = \sup_{F \in \mathcal{F}_0} \mathbf{P}_F\left(\varphi_n^{**} = 1\right) \to 0, \quad n \to \infty.$$

Next we evaluate the error probability of the second type .

Lemma 4.5 *Let $F \in \mathcal{F}_1$. Then for n large enough*

$$\langle F - F_{\theta_n} \rangle \geq c_n/2 \, .$$

Proof. Let $F \in \mathcal{F}_1$ be fixed and

$$\theta_F = \underset{\theta \in \Theta}{\operatorname{arginf}} \| F - F_\theta \| \, .$$

By the triangle inequality and Lemma 4.4 one has

$$\langle F - F_{\theta_F} \rangle \leq \langle F - F_{\theta_n} \rangle + \langle F_{\theta_n} - F_{\theta_F} \rangle \leq \langle F - F_{\theta_n} \rangle + C_\theta^2 \frac{\ln^2 n}{n}.$$

It remains to check that the inequality $\|F - F_{\theta_F}\| \geq c_n$ implies $\langle F - F_{\theta_F} \rangle \geq c_n/2$. For n large enough that is obviously the case.

The following Lemma is a direct consequence of assumptions $(E3)$ and (D).

Lemma 4.6 *There exist constants C_π , σ^* and V^* such that*

$$|\Pi_\beta(X_i) K_{h,\beta}(X_i - X_j)| \leq C_\pi \, |\Pi_\beta(X_j) K_{h,\beta}(X_i - X_j)| \qquad \forall \beta, X_i, X_j \tag{35}$$

$$\sup_i \sigma_i \leq \sigma^*. \tag{36}$$

and

$$\sup_\beta V_\beta \leq V^*. \tag{37}$$

Recall now that each function $F(\cdot)$ from $\mathcal{F}_1$ is of the form $F(x) = f(x^\top \beta_0)$ with some $\beta_0 \in S_d$. As a consequence $F(\cdot)$ should be well approximated by the smoothing operator $\mathcal{K}_\beta$ with β coinciding or close to β_0. More precise, the following can be stated.

Lemma 4.7 *There is a positive constant C_b such that for each $F(\cdot) \in \mathcal{F}_1$, $F(x) = f(x^\top \beta_0)$,*

$$\langle B_{\beta_n} \rangle \leq C_b h^{2s} \tag{38}$$

with

$$\beta_n = \underset{\beta \in S_{n,d}}{\mathit{arginf}} \, |\beta - \beta_0| \, . \tag{39}$$

Proof. The definition of the grid $S_{n,d}$ provides $|\beta_n - \beta_0| \leq h^{2s+2}$. Then, it is well known, e.g. from Ibragimov and Khasminskii (1981), that for $F(x) = f(x^\top \beta_0)$ with $f \in \Sigma(s, L)$ one has

$$\langle B_{\beta_0} \rangle = \langle \mathcal{K}_{\beta_0} F - F \rangle \leq L' h^{2s+1} \tag{40}$$

with $L' = L\|K\|/(s-1)!$ But

$$
\begin{aligned}
|\langle B_{\beta_n}\rangle - \langle B_{\beta_0}\rangle| &\le \langle B_{\beta_n} - B_{\beta_0}\rangle \le \\
&\le \langle \mathcal{K}_{\beta_0} F - \mathcal{K}_{\beta_n} F\rangle \le \\
&\le \frac{1}{n}\sum_i \Bigg| \Pi_{\beta_n}(X_i)\sum_{j\ne i} F(X_j)\, K_{h,\beta_n}(X_i - X_j) - \\
&\qquad - \Pi_{\beta_0}(X_i)\sum_{j\ne i} F(X_j)\, K_{h,\beta_0}(X_i - X_j)\Bigg|.
\end{aligned}
$$

Now using assumptions (D) and $(K1)-(K5)$ we obtain

$$
\begin{aligned}
\left|\Pi^{-1}_{\beta_n}(X_i) - \Pi^{-1}_{\beta_0}(X_i)\right| &\le \sum_{j\ne i} |K_{h,\beta_n}(X_i - X_j) - K_{h,\beta_0}(X_i - X_j)| \le \\
&\le C\Pi_{\beta_0}(X_i)\frac{|\beta_n - \beta_0|}{h} \qquad (41)
\end{aligned}
$$

and similarly

$$
\sum_{j\ne i} |F(X_j)K_{h,\beta_n}(X_i - X_j) - F(X_j)K_{h,\beta_0}(X_i - X_j)| \le C\Pi_{\beta_0}(X_i)\frac{|\beta_n - \beta_0|}{h}. \qquad (42)
$$

Putting together (41) and (42) we conclude that

$$
|\langle B_{\beta_n}\rangle - \langle B_{\beta_0}\rangle| \le C\frac{|\beta_n - \beta_0|}{h} \le Ch^{2s+1}
$$

and the lemma follows with $C_b = L' + 1$.

To complete the proof for the tests φ_n^{**} it remains to note that for each $F \in \mathcal{F}_1$

$$
T_n^{**} \ge \frac{n\sqrt{h}}{V_{\beta n}}\,|\langle F - F_{\theta_n}\rangle - \langle B_{\beta_n}\rangle| + \frac{\xi_{\beta_n}}{V_{\beta_n}}
$$

and that if

$$
\langle F - F_{\theta_n}\rangle \ge C_b h^{2s} + \frac{2V^*}{n\sqrt{h}}\sqrt{(2+\delta)\ln N}, \qquad (43)
$$

with V^* from Lemma 4.6, then by Lemma 4.3 we obtain

$$
\begin{aligned}
&\mathbf{P}\left(T_n^{**} < \sqrt{(2+\delta)\ln N}\right) \le \\
&\le \mathbf{P}\left(\frac{n\sqrt{h}}{V_{\beta n}}\frac{2V^*}{n\sqrt{h}}\sqrt{(2+\delta)\ln N} + \frac{\xi_{\beta_n}}{V_{\beta_n}} < \sqrt{(2+\delta)\ln N}\right) \le \\
&\le \mathbf{P}\left(\left|\frac{\xi_{\beta_n}}{V_{\beta_n}}\right| > \sqrt{(2+\delta)\ln N}\right) \to 0, \quad n\to\infty.
\end{aligned}
$$

Finally we remark that $\ln N \le C\ln n$ and the choice of h by (8) yields

$$
C_b h^{2s} + \frac{2V^*}{n\sqrt{h}}\sqrt{(2+\delta)\ln N} \le C'\left(\frac{\sqrt{\ln n}}{n}\right)^{\frac{4s}{4s+1}} = C'h^{2s}
$$

i.e. (43) holds true if c_n in the definition of the alternative H_1 is taken with $c_n^2 \geq 2C'h^{2s}$. This completes the proof for the tests φ_n^{**}

Now we explain why the statistics T_n^{**} can be considered in place of T_n^*. The idea is to show that the difference $T_n^{**} - T_n^*$ is relatively small (being compared with the test level $\sqrt{2\ln N}$ or deviation $\langle F - F_{\theta_n}\rangle$). First we treat the preliminary parametric estimator $\tilde{\theta}_n$. Denote for given $F \in \mathcal{F}_0 \cup \mathcal{F}_1$

$$d_n(F) = \langle F - F_{\theta_n}\rangle + \frac{\ln^2 n}{n},$$

θ_n being from (27)

Lemma 4.8 *Uniformly in $F \in \mathcal{F}_0 \cup \mathcal{F}_1$ we have for each $\delta > 0$*

$$\mathbf{P}\left(\frac{1}{d_n(F)}\left|\left\langle F - \tilde{F}_0\right\rangle - \langle F - F_{\theta_n}\rangle\right| > \delta\right) \to 0, \tag{44}$$

$$\mathbf{P}\left(\frac{1}{d_n(F)}\left|\left\langle F - \tilde{F}_0, \varepsilon\right\rangle\right| > \delta\right) \to 0.$$

Proof. Let us fix some $\delta > 0$ and some $\theta \in \Theta_n$. First we show that the probability of the event

$$\left\{|\langle F - F_\theta, \varepsilon\rangle| > \delta\left(\langle F - F_\theta\rangle + \frac{\ln^2 n}{n}\right)\right\}$$

is asymptotically small. More precise, we state the following assertion:

$$\sum_{\theta\in\Theta_n} \mathbf{P}\left(|\langle F - F_\theta, \varepsilon\rangle| > \delta\left(\langle F - F_\theta\rangle + \frac{\ln^2 n}{n}\right)\right) \to 0, \quad n \to \infty. \tag{45}$$

In fact, if we put $d_\theta^2 = \mathbf{E}\,|\langle F - F_\theta, \varepsilon\rangle|^2$ then we have

$$\begin{aligned} d_\theta^2 &= \mathbf{E}\left|\frac{1}{n}\sum_i \varepsilon_i \left[F(X_i) - F_\theta(X_i)\right]\right|^2 = \\ &= \frac{1}{n^2}\sum_i \sigma_i^2 \left|F(X_i) - F_\theta(X_i)\right|^2. \end{aligned}$$

Using Lemma 4.6 we have

$$d_\theta^2 \leq \frac{\sigma^{*2}}{n^2}\sum_i |F(X_i) - F_\theta(X_i)|^2 = \frac{\sigma^{*2}}{n}\langle F - F_\theta\rangle.$$

Further,

$$\frac{1}{2}\left(\langle F - F_\theta\rangle + \frac{\ln^2 n}{n}\right) \geq \sqrt{\langle F - F_\theta\rangle \frac{\ln^2 n}{n}}$$

and

$$\begin{aligned}
&\mathbf{P}\left(|\langle F - F_\theta, \varepsilon\rangle| > \delta\left(\langle F - F_\theta\rangle + \frac{\ln^2 n}{n}\right)\right) \le \\
&\le \ \mathbf{P}\left(\frac{1}{d_\theta}|\langle F - F_\theta, \varepsilon\rangle| > \frac{\delta}{d_\theta}\ln n\sqrt{\langle F - F_\theta\rangle / n}\right) \le \\
&\le \ \mathbf{P}\left(\frac{1}{d_\theta}|\langle F - F_\theta, \varepsilon\rangle| > \frac{\delta}{\sigma^*}\ln n\right).
\end{aligned}$$

Now we use an estimate of the large deviation probability for the centered and normalized random variables $\frac{1}{d_\theta}\langle F - F_\theta, \varepsilon\rangle$, see Lemma 4.11 below. Indeed, for n large enough

$$\begin{aligned}
&\sum_{\theta\in\Theta_n}\mathbf{P}\left(\frac{1}{d_\theta}\langle F - F_\theta, \varepsilon\rangle > \frac{\delta}{\sigma^*}\ln n\right) \le \\
&\le \ \sum_{\theta\in\Theta_n}\exp\left\{-\frac{\delta^2}{2\sigma^{*2}}\ln^2 n\right\} \le n^d \exp\left\{-(d+1)\ln n\right\} \le n^{-1}
\end{aligned}$$

which implies (45). Here we used that the cardinality of Θ_n is of order n^d. Let $\theta \in \Theta_n$ be such that

(46) $$\langle F - F_\theta\rangle - \langle F - F_{\theta_n}\rangle > 2\delta d_n(F).$$

For δ small enough this yields

(47) $$\langle F - F_\theta\rangle - \langle F - F_{\theta_n}\rangle > \delta\left(\langle F - F_\theta\rangle + \langle F - F_{\theta_n}\rangle\right).$$

Now by definition of $\tilde\theta_n$ we obtain through (46) and (47)

$$\begin{aligned}
\left\{\tilde\theta_n = \theta\right\} \ &\subseteq \ \left\{\langle Y - F_\theta\rangle \le \langle Y - F_{\theta_n}\rangle\right\} = \\
&= \ \left\{\langle F - F_\theta + \varepsilon\rangle \le \langle F - F_{\theta_n} + \varepsilon\rangle\right\} = \\
&= \ \left\{\langle F - F_\theta\rangle - \langle F - F_{\theta_n}\rangle \le 2\langle F - F_\theta, \varepsilon\rangle + 2\langle F - F_{\theta_n}, \varepsilon\rangle\right\} \subseteq \\
&\subseteq \ \left\{\langle F - F_\theta, \varepsilon\rangle > \frac{\delta}{2}\langle F - F_\theta\rangle\right\} \cup \left\{\langle F - F_{\theta_n}, \varepsilon\rangle > \frac{\delta}{2}\langle F - F_{\theta_n}\rangle\right\}.
\end{aligned}$$

Using this relation and (45) we deduce

$$\begin{aligned}
&\mathbf{P}\left(\left|\left\langle F - F_{\tilde\theta_n}\right\rangle - \langle F - F_{\theta_n}\rangle\right| > 2\delta d_n(F)\right) \le \\
&\le \ \sum_{\theta\in\Theta_n}\mathbf{1}\left(|\langle F - F_\theta\rangle - \langle F - F_{\theta_n}\rangle| > 2\delta d_n(F)\right)\mathbf{P}\left(\tilde\theta_n = \theta\right) \le \\
&\le \ \sum_{\theta\in\Theta_n}\mathbf{P}\left(\langle F - F_\theta, \varepsilon\rangle > \frac{\delta}{2}\langle F - F_\theta\rangle\right) \to 0, \quad n \to \infty,
\end{aligned}$$

that proves (44). The second statement of the lemma follows directly from (45).

The next step is to show that the last two terms in the expansion (29) are vanishing.

Lemma 4.9 *Given F let*

$$b_\beta = \langle B_\beta \rangle + \frac{\ln^2 n}{n}.$$

Then uniformly in $F \in \mathcal{F}_0 \cup \mathcal{F}_1$ for each $\delta > 0$ the following assertions hold:

$$\begin{aligned} \sum_{\beta \in S_{n,d}} \mathbf{P}\left(\langle B_\beta, \varepsilon \rangle > \delta b_\beta\right) &\to 0, \\ \sum_{\beta \in S_{n,d}} \mathbf{P}\left(\langle B_\beta, \mathcal{K}_\beta \varepsilon \rangle > \delta b_\beta\right) &\to 0. \end{aligned}$$

Remark **4.1** The statements of this lemma yield immediately that

$$\mathbf{P}\left(\langle B_\beta, \varepsilon \rangle \leq \delta b_\beta\,, \quad \forall \beta \in S_{n,d}\right) \to 1$$

and similarly for $\langle B_\beta, \mathcal{K}_\beta \varepsilon \rangle$.

Proof. The statements of the lemma are proved in the same manner as in the last part of the proof of Lemma 4.8. For the second statement we use in addition the fact that

$$\operatorname{Var} \langle B_\beta, \mathcal{K}_\beta \varepsilon \rangle \leq \frac{C}{n} \langle B_\beta \rangle\,. \tag{48}$$

Indeed, using assumptions $(E1)$-$(E3)$ and $(K1)$-$(K5)$, Lemma 4.6 and Jensen's inequality we have

$$\begin{aligned} \mathbf{E}\left|\langle B_\beta, \mathcal{K}_\beta \varepsilon \rangle\right|^2 &= \frac{1}{n^2} \mathbf{E} \left| \sum_i B_\beta(X_i) \Pi_\beta(X_i) \sum_{j \neq i} \varepsilon_j K_{h,\beta}(X_i - X_j) \right|^2 = \\ &= \frac{1}{n^2} \mathbf{E} \left| \sum_j \varepsilon_j \sum_{i \neq j} B_\beta(X_i) \Pi_\beta(X_i) K_{h,\beta}(X_i - X_j) \right|^2 = \\ &= \frac{1}{n^2} \sum_j \sigma_j^2 \left| \sum_{i \neq j} B_\beta(X_i) \Pi_\beta(X_i) K_{h,\beta}(X_i - X_j) \right|^2 \leq \\ &\leq \frac{1}{n^2} \sigma^{*2} C_\pi^2 \sum_j \Pi_\beta^2(X_j) \left| \sum_{i \neq j} B_\beta(X_i) K_{h,\beta}(X_i - X_j) \right|^2 \leq \\ &\leq \frac{1}{n^2} \sigma^{*2} C_\pi^2 \sum_j \left| \frac{\sum_{i \neq j} B_\beta(X_i) K_{h,\beta}(X_i - X_j)}{\sum_{i \neq j} K_{h,\beta}(X_i - X_j)} \right|^2 \leq \\ &\leq \frac{1}{n^2} \sigma^{*2} C_\pi^2 C \langle B_\beta \rangle\,. \end{aligned}$$

Next we show that the quantities $\tilde{E}_\beta$ and $\tilde{V}_\beta$ estimate E_β and V_β good enough.

Lemma 4.10 *For each $\delta > 0$ and uniformly in $F \in \mathcal{F}_0 \cup \mathcal{F}_1$*

$$\mathbf{P}\left(\sup_{\beta \in S_{n,d}} \left|\tilde{E}_\beta - E_\beta\right| > \frac{1}{n\sqrt{h}\ln n}\right) \to 0,$$

$$\mathbf{P}\left(\sup_{\beta \in S_{n,d}} \left|\frac{\tilde{V}_\beta}{V_\beta} - 1\right| > \delta\right) \to 0.$$

Proof. The assumption $(E3)$ implies for each $j = 1, \ldots, n$

$$\left|\sigma_j^2 - \tilde{\sigma}_j^2\right| \le C_\sigma \left|\tilde{F}_1(X_j) - F(X_j)\right|$$

and hence

$$\left|\tilde{E}_\beta - E_\beta\right| \le \frac{1}{n} \sum_i \sum_{j \ne i} \left|\sigma_j^2 - \tilde{\sigma}_j^2\right| \Pi_\beta^2(X_i) K_{h,\beta}^2(X_i - X_j).$$

Now by the design and kernel properties we derive for each $j = 1, \ldots, n$

$$\sum_{j \ne i} \Pi_\beta^2(X_i) K_{h,\beta}^2(X_i - X_j) \le \frac{C}{nh}$$

and using Cauchy-Schwarz inequality we obtain

$$\left|\tilde{E}_\beta - E_\beta\right| \le \frac{C}{n^2 h} \sum_i \left|\tilde{F}_1(X_j) - F(X_j)\right| \le \frac{C}{n^2 h}\left[\frac{1}{n}\sum_i \left|\tilde{F}_1(X_j) - F(X_j)\right|^2\right]^{1/2}.$$

The pilot estimator $\tilde{F}_1$ fulfills with high probability

$$\left\langle \tilde{F}_1 - F \right\rangle \le C n^{-\frac{2s}{2s+d}}.$$

Hence using the inequality $\frac{2s}{2s+d} > \frac{1}{4s+1}$ and the definition of h we arrive to the conclusion that

$$n\sqrt{h}\left|\tilde{E}_\beta - E_\beta\right| \le \frac{C}{\sqrt{h}} n^{-\frac{s}{2s+d}} = o\left(\frac{1}{\ln n}\right).$$

Lemmas 4.8–4.10 together imply the asymptotic equivalence of the tests based on T_β and T'_β. We finish the proof of the theorem with a result on probabilities of deviations of centered and normalized sums of independent errors ε_i over the logarithmic level. The following lemma was already used in the proof of Lemma 4.8.

Lemma 4.11 *For each positive constants r, a the following relation holds uniformly in functions F from the Lipschitz class $\Sigma_d(1, L)$ of functions in $\mathbb{R}^d$:*

$$n^r \mathbf{P}\left(\xi(F) > a \ln n\right) \to 0, \quad n \to \infty,$$

where

$$\xi(F) = \frac{\langle F, \varepsilon \rangle}{\sqrt{\mathbf{E}\langle F, \varepsilon\rangle^2}}.$$

Proof. We proceed in a standard way using the exponential inequality and boundedness of errors ε_i due to $(E1)$. The details are omitted.

5 A simulation and an application

The purpose of our simulation experiments was to study the quantiles of the test statistic T_n^* and the power of the test in finite samples. All calculations have been performed in the languages GAUSS and XploRe (Härdle, Klinke and Turlach (1995)). The observations were generated according to a binary response model. The explanatory variables were identically independent uniform distributed on $[-1, 1]$. We took the parameter $\theta = \binom{1}{1}\frac{1}{\sqrt{2}}$ and considered the functions

$$f_0(u) = \frac{1}{1 + \exp^{-u}} \tag{49}$$

$$f_1(u) = f_0(u) + \eta \cdot \varphi'(u) \tag{50}$$

$$f_2(u) = 1 - \exp(-\exp(u)) \tag{51}$$

for different $0 < \eta \le 1$, where φ is the density function of the standard normal distribution. While f_0 is a logit function, f_1 consists of a logit disturbed by a bump (figure 3). The response Y under H_0 was generated such that $P(Y = 1|x^T\theta_0 = u) = f_0(u)$. We are thus interested in the hypothesis H_0

$$H_0 : F_\theta(x) = \mathbf{E}[Y|u(x,\theta) = u] = f_0(u) \qquad , \ \theta \in S_2$$

In a first step we calculated empirically the 90 and 95 percent quantiles of T_n^* for $n = 100$ and 200 observations generated by f_0. They were used then as rejection boundaries, defined as $\sqrt{(2+\delta)\ln N}$, see (24). We calculated T_n^* by optimizing T_β over a grid, see (23), with $N = 50$ gridpoints. As kernel function K we used always the quartic kernel

$$K(u) = \frac{15}{16}(1 - u^2)^2 \mathbf{1}_{\{|u|<1\}}$$

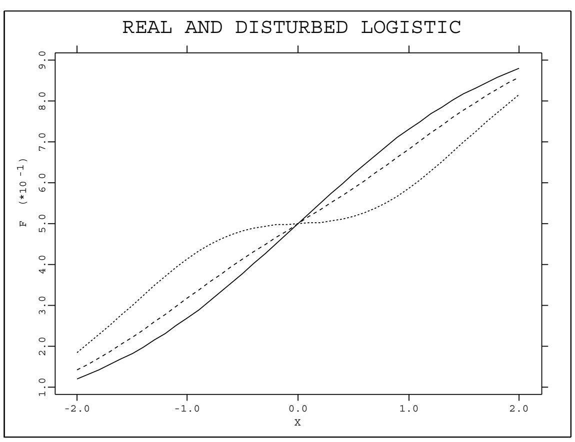

Figure 3: *solid line:* f_0, *dashed line:* f_1 with $\eta = 0.2$, *pointed line:* f_1 with $\eta = 0.6$

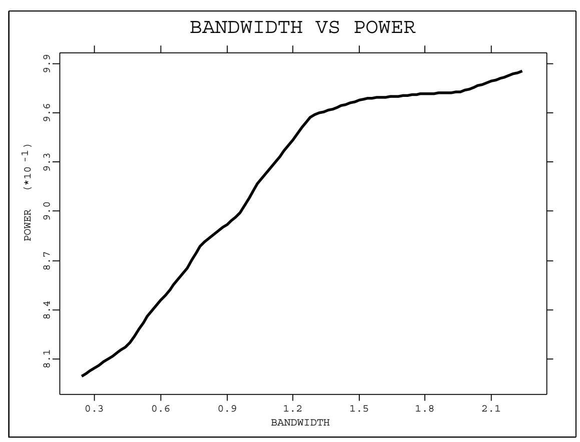

Figure 4: Power function of the test with respect to the bandwidth for funtion f_{1c}

In the second step we analyzed the effect of increasing sample size on the power. In table 1 we show the power of the test when the data were generated with functions f_{1a}, that is f_1 for $\eta = 0.2$, f_{1c}, where $\eta = 0.6$ and f_2. In order not to oversmooth we used the bandwidth $h_1 = h = 0.5$ for $n = 100,\ 200$ and $h_1 = h = 0.25$ for $n = 350,\ 500$. Although we substituted for speed reasons in the cases $n = 350$ and 500 $\tilde{V}_\beta$ by $\tilde{V}_{\hat{\theta}}$ for all β, the power increases very fast with n. Therefore, it could be of interest to compare the power with regard to the bump η in the logit model. In table 2 we show for $n = 200$ and 350 the power of the test as a function of η. We see that for $\eta > 0.4$ this test procedure works very well.

Table 1: *Power and rejection boundaries for different alternatives.*

$n\ ,\ h\ =$	100 , 0.5		200 , 0.5		350 , 0.25		500 , 0.25	
level	5%	10%	5%	10%	5%	10%	5%	10%
rejection boundary	4.00	3.35	3.30	3.25	3.75	2.90	3.20	2.76
f_{1a}	0.056	0.096	0.112	0.215	0.133	0.207	0.150	0.200
f_{1c}	0.224	0.294	0.530	0.690	0.798	0.856	0.900	0.960
f_2	0.316	0.376	0.946	0.991	0.995	1.000	0.995	1.000

Table 2: *Power for different bumps η .*

	$\eta =$	0.2		0.4		0.6		1.0	
	level	5%	10%	5%	10%	5%	10%	5%	10%
$n,\ h$	200, 0.50	0.112	0.215	0.227	0.419	0.530	0.690	0.687	0.801
	350, 0.25	0.133	0.207	0.321	0.478	0.798	0.856	0.889	0.926

The last step of the simulation experiment was the study of bandwidth choice. For the sake of simplicity we set $h_1 = h$ as above. First we always have had to determine numerically the rejection boundaries for the special bandwidth h. Here we observed shrinking boundaries, when h grew from 0.25 up to 2.25 . In figure 4 we plot the bandwidth vs the power of the test with observations generated by f_{1c}. Obviously for this kind of alternative we get better power for larger bandwidths.

In the introductory example we dealt with youth unemployement. The question is, can we explain the youth unemployement with the aforementioned predictor variables X in a single index model with logit link? In the application of this dataset, we used a sligthly modified procedure as described in Proenca and Ritter (1995). Further we rescaled the explanatory variables of each dimension to $[-1, 1]$. Since there are three dimensions $(d = 3)$ for a sample size of $n = 462$, we chose the bandwidth h_1 large, definitively 1.5, whereas $h = 0.3$. By Monte Carlo studies described above we determined the 90 and 95% one side quantiles of T^*_{462} and got 1.74 respectively 2.38 . Now we ran the test for our data and got the statistic value $T^*_{462} = 3.076$ for $\beta = (-0.18010, -0.10725, 0.97778)$. For purpose of comparison in table 3 we switch the norm of β and set his first component equal to the corresponding one of θ, the parameter of the logit fit in figure 1.

Table 3: *Comparison of θ and β.*

explanatory variables	intercept	earnings as an apprentice	percentage of apprentices divided by employees	unemployed rate
θ	-2.40996	-0.07999	-0.17989	0.95113
β	-	-0.07999	-0.04763	0.43422

Appendix

Proof of Theorem 1. To simplify our exposition and to emphasize the main idea we consider the case when the parametric family consists of one point, namely, a zero regression function, and errors ε_i are independent and standard Gaussian. Moreover, we assume random design with a design density $\pi(x)$ in $I\!R^d$ of the form $\pi(x) = \pi_1(|x|)$ where a univariate function $\pi_1(\cdot)$ is compactly supported on $[-1, 1]$, symmetric, twice continuously differentiable and satisfies $\pi_1(t) = 3/4$ for $|t| \le 1/2$.

The idea of the proof is standard. We replace the minimax problem by a Bayes one where we consider instead of the set $\mathcal{F}_1$ of alternatives one Bayes alternative corresponding to a prior ν concentrated on $\mathcal{F}_1$. We try to choose this prior ν in such a way that the likelihood $Z_\nu = d\mathbf{P}_\nu/d\mathbf{P}_0$ is close to 1 where the measure $\mathbf{P}_\nu$ is the Bayes measure for the prior ν and $\mathbf{P}_0$ corresponds to the case of zero regression function. The Neyman-Pearson Lemma yields that the hypothesis $H_0 : \mathbf{P} = \mathbf{P}_0$ can not be consistently distinguished versus the Bayes alternative $H_\nu : \mathbf{P} = P_\nu$ and hence versus the composite alternative $H_1 : \mathbf{P} \in \mathcal{F}_1$.

Now we describe the structure of the prior ν. Let $g(\cdot)$ be some function from the Hölder class $\Sigma(s, L)$, supported on $[-1, 1]$ and satisfying the conditions

$$\int g(t)\,dt = 0, \qquad \|g\|^2 = \int g^2(t)\,dt > 0. \tag{52}$$

Set

$$h = \left(\frac{a\sqrt{\ln n}}{n}\right)^{\frac{2}{4s+1}} \tag{53}$$

where a constant a will be chosen later. Denote by $\mathcal{I}_n$ the partition of the interval $\left[-\frac{1}{2}, \frac{1}{2}\right]$ into intervals of length h. Without loss of generality we assume that the cardinality of the set $\mathcal{I}_n$ coincides with $1/h$

$$\#\mathcal{I}_n = \frac{1}{h}. \tag{54}$$

For each interval $I \in \mathcal{I}_n$ introduce a function $g_I(t)$ of the form

$$g_I(t) = h^s g\left(\frac{t - t_I}{h}\right), \tag{55}$$

t_I being the center of I. Evidently $g_I(\cdot)$ is supported on I, $g_I \in \Sigma(s, L)$ and the followings hold for h small enough:

$$\int g_I(t)\,dt = 0, \qquad \int g_I^2(t)\,dt = h^{2s+1}\|g\|^2. \tag{56}$$

Let now μ be a set of binary values $\{\mu_I, I \in \mathcal{I}_n\}$ i.e. $\mu_I = \pm 1$. Define a function $G_\mu(t)$ with

$$G_\mu(t) = \sum_{I \in \mathcal{I}_n} \mu_I \, g_I(t). \tag{57}$$

This function $G_\mu \in \Sigma(s, L)$ vanishes outside $\left[-\frac{1}{2}, \frac{1}{2}\right]$ and by (56)

$$\int G_I^2(t)\,dt = \sum_{I \in \mathcal{I}_n} \int g_I^2(t)\,dt = \frac{1}{h} h^{2s+1} \int g^2(t)\,dt = h^{2s} \|g\|^2 . \tag{58}$$

Taking into account (53) we see that the distance between zero function and each G_μ is just of the rate c_n^2 from Theorems 2.1 and 2.2.

Denote by $\mathcal{M}_n$ the set of all possible collections $\{\mu_I, I \in \mathcal{I}_n\}$ with binaries $\mu_I = \pm 1$, and let $m(d\mu)$ be the uniform measure on $\mathcal{M}_n$. This measure can be represented as the direct product of binary measures $m_I(d\mu_I)$ with $m_I(\mu_I = \pm 1) = 1/2$.

Now we pass to the semiparametric model. Let S_n be a grid on the unit sphere S_d with the step b_n,

$$b_n = h^{1/8}, \tag{59}$$

h being from (53). This means that $|\beta - \beta'| \geq b_n = h^{1/8}$ for each $\beta, \beta' \in S_n$, $\beta \neq \beta'$. Below we will use that for some $\alpha > 0$

$$N = \# S_n \asymp n^\alpha \tag{60}$$

and for n large enough

$$\frac{h \ln n}{|\beta - \beta'|^4} \leq \frac{h \ln n}{b_n^4} \leq h^{1/4} \quad \forall \beta, \beta' \in S_n,\, \beta \neq \beta'. \tag{61}$$

For each $\beta \in S_n$ and each $\mu \in \mathcal{M}_n$ define the multivariate function $G_{\beta,\mu}(x)$ on $\mathbb{R}^d$ with

$$G_{\beta,\mu}(x) = G_\mu(x^\top \beta).$$

It is clear that the function $G_{\beta,\mu}(x)$ is Hölder, $G_{\beta,\mu}(x) \in \Sigma_d(s, L)$, and by (58) we get

$$\int G_{\beta,\mu}^2(x)\, \pi(x)\,dx = \int G_\mu^2(x^\top \beta)\, \pi_1(|x|)\,dx = \int G_\mu^2(t)\, \pi_2(t)\,dt = C_0 h^{2s} \tag{62}$$

with $\pi_2(t) = \frac{d}{dt} \int \mathbf{1}(x^\top \beta \leq t)\, \pi_1(|x|)\,dx$ and $C_0 \in \left[\frac{1}{2}\|g\|^2, \|g\|^2\right]$.

Finally we take the prior ν as the uniform measure on the set of functions $\{G_{\beta,\mu}\}$, $\beta \in S_n$, $\mu \in \mathcal{M}_n$, and

$$\mathbf{P}_\nu = \frac{1}{N} \sum_{\beta \in S_n} \frac{1}{M} \sum_{\mu \in \mathcal{M}_n} \mathbf{P}_{G_{\beta,\mu}} . \tag{63}$$

Here $M = \#\mathcal{M}_n = 2^{1/h}$, N being from (60). Denote also $Z_\nu = \frac{d\mathbf{P}_\nu}{d\mathbf{P}_0}$ and notice that this likelihood can be represented in the form $Z_\nu = \frac{1}{N}\sum_{\beta\in S_n} Z_\beta$ with

$$Z_\beta = \frac{1}{M}\sum_{\mu\in\mathcal{M}_n} Z_{\beta,\mu} = \frac{1}{M}\sum_{\mu\in\mathcal{M}_n} d\mathbf{P}_{G_{\beta,\mu}}/d\mathbf{P}_0. \tag{64}$$

Our goal is to prove that for a small enough in (53) one has

$$Z_\nu \to 1 \tag{65}$$

under the measure $\mathbf{P}_0$.

We start from a decomposition and an asymptotic expansion for each Z_β from (64). For that we need some more notation. Fix some $\beta \in S_n$ and put

$$\sigma^2_{\beta,I} = \sum_i g_I^2(X_i^\top \beta), \quad I \in \mathcal{I}_n\,, \tag{66}$$

$$\xi_{\beta,I} = \frac{1}{\sigma_{\beta,I}}\sum_i g_I(X_i^\top\beta)\,\varepsilon_I, \quad I\in\mathcal{I}_n\,. \tag{67}$$

We see that $\xi_{\beta,I}$ are standard normal and independent for different $I \in \mathcal{I}_n$, and

$$\sum_i G^2_{\beta,\mu}(X_i) = \sum_i G^2_\mu(X_i^\top\beta) = \sum_{I\in\mathcal{I}_n}\sigma^2_{\beta,I}\,.$$

Recall that we assume the random design and

$$\mathbf{E}\sum_i G^2_{\beta,\mu}(X_i) = n\int G^2_{\beta,\mu}(x)\,\pi(x)\,dx = nC_0h^{2s}. \tag{68}$$

Similarly for each $\sigma^2_{\beta,I}$

$$\mathbf{E}\sigma^2_{\beta,I} = n\int g_I^2(x^\top\beta)\,\pi(x)\,dx = n\int g_I^2(x^\top\beta)\,\pi_1(|x|)\,dx = nC_Ih^{2s+1} \tag{69}$$

where C_I does not depends on β and $C_I \in \left[C_0/\sqrt{2}, \sqrt{2}C_0\right]$.

Lemma 5.1

$$Z_\beta = \prod_{I\in\mathcal{I}_n} ch(\sigma_{\beta,I}\,\xi_{\beta,I})e^{-\frac{1}{2}\sigma^2_{\beta,I}}$$

where $ch(z) = \frac{1}{2}\left(e^z + e^{-z}\right)$.

Proof. By Girsanov formulae and (66)-(67)

$$\begin{aligned}
Z_{\beta,\mu} &= \exp\left\{\sum_i G_{\beta,\mu}(X_i)\,\varepsilon_i - \frac{1}{2}\sum_i G^2_{\beta,\mu}(X_i)\right\} = \\
&= \exp\left\{\sum_{I\in\mathcal{I}_n}\mu_I\sigma_{\beta,I}\xi_{\beta,I} - \frac{1}{2}\sum_{I\in\mathcal{I}_n}\sigma^2_{\beta,I}\right\} = \\
&= \prod_{I\in\mathcal{I}_n}\exp\left\{\mu_I\sigma_{\beta,I}\xi_{\beta,I} - \frac{1}{2}\sigma^2_{\beta,I}\right\}.
\end{aligned}$$

Now the lemma's assertion follows from the direct product structure of the measure $m(d\mu)$.

Denote also

(70) $$v_\beta^2 = \frac{1}{2} \sum_{I \in \mathcal{I}_n} \sigma_{\beta,I}^4 ,$$

(71) $$\zeta_\beta = \frac{1}{v_\beta} \sum_{I \in \mathcal{I}_n} \sigma_{\beta,I}^2 \left(\xi_{\beta,I}^2 - 1 \right) .$$

Lemma 5.2 *The following statements hold:*

(i) $\mathbf{E}\zeta_\beta = 0;$

(ii) $\mathbf{E}\zeta_\beta^2 = 1;$

(iii) $v_\beta^2 = C_1 n^2 h^{4s+1} = C_1 \ln n \quad$ *with* $C_1 \le a \quad ;$

(iv) There exists an independent standard normal r.v. $\tilde{\zeta}_\beta$ *that*

$$\ln n \sup_{\beta \in S_n} \mathbf{E}_0 \left(\tilde{\zeta}_\beta - \zeta_\beta \right)^2 \to 0.$$

Proof. The first two statements are obvious. (iii) follows from (69). Finally, (iv) is the application of the Strassen type invariance principle (see, e.g. ??).

The next step is the asymptotic expansion for each Z_β.

Lemma 5.3 *The following statements are satisfied uniformly in* $\beta \in S_n$*: for each* $\delta > 0$

(i)

$$\mathbf{P}_0 \left(\left| Z_\beta - \exp\left\{ v_\beta \zeta_\beta - \frac{1}{2} v_\beta^2 \right\} \right| > \delta \right) \to 0;$$

(ii)

$$\mathbf{P}_0 \left(\left| \tilde{Z}_\beta - \exp\left\{ v_\beta \tilde{\zeta}_\beta - \frac{1}{2} v_\beta^2 \right\} \right| > \delta \right) \to 0;$$

Proof. The first statement is equivalent to the following one:

$$\mathbf{P}_0 \left(\left| \ln Z_\beta - v_\beta \zeta_\beta + \frac{1}{2} v_\beta^2 \right| > \delta \right) \to 0.$$

But the latter can be obtained using Taylor expansion for $\ln Z_\beta$

$$\begin{aligned} \ln Z_\beta &= \sum_{I \in \mathcal{I}_n} \ln ch(\sigma_{\beta,I} \xi_{\beta,I}) - \frac{1}{2} \sigma_{\beta,I}^2 = \\ &= \sum_{I \in \mathcal{I}_n} \left[\frac{1}{2} \sigma_{\beta,I}^2 \left(\xi_{\beta,I}^2 - 1 \right) - \frac{1}{12} \sigma_{\beta,I}^4 \xi_{\beta,I}^4 + O(\sigma_{\beta,I}^6 \xi_{\beta,I}^6) \right] \end{aligned}$$

and the following asymptotic relations which hold uniformly in β

$$\mathbf{P}_0 \left(\left| \sum_{I \in \mathcal{I}_n} \sigma^4_{\beta,I} (\xi^4_{\beta,I} - 3) \right| > \delta \right) \;\; \to \;\; 0;$$
$$\mathbf{P}_0 \left(\left| \sum_{I \in \mathcal{I}_n} \sigma^6_{\beta,I} \xi^6_{\beta,I} \right| > \delta \right) \;\; \to \;\; 0;$$

for details we refer to Ingster(1993).

The second statement of the lemma follows directly from (iii) of Lemma 5.2.

Now we arrive at the central point of the proof. Actually we prove that "submodels" corresponding to different β are in some sense asymptotically independent. That is why we have to pay with the extra log-term for the choice of "direction" β.

Lemma 5.4 *There exist a universal constant R such that for any $\beta, \beta' \in S_n$, $\beta \neq \beta'$,*

$$|\mathbf{E}\, \zeta_\beta \zeta_{\beta'}| \le \frac{Rh}{|\beta - \beta'|^4}. \tag{72}$$

Proof. Let us fix some β, β' from S_n. Denote by ρ their scalar product,

$$\rho = (\beta, \beta').$$

Now fix also some I, I' from $\mathcal{I}_n$ and set

$$r = r(\beta, I, \beta', I') = \mathbf{E}\, \xi_{\beta,I} \xi_{\beta',I'}\,.$$

Using normality of $\xi_{\beta,I}$ and $\xi_{\beta',I'}$ we calculate easily

$$\mathbf{E} \left(\xi^2_{\beta,I} - 1 \right) \left(\xi^2_{\beta',I'} - 1 \right) = 4r^2 - 2r. \tag{73}$$

Below we state that r satisfies the condition

$$|r| \le Ch^2/(1-\rho)^2 \tag{74}$$

with some universal constant C and now we show that this implies (72). In fact, through (73) one has

$$\begin{aligned} \mathbf{E}\, \zeta_\beta \zeta_{\beta'} &= \mathbf{E} \frac{1}{v_\beta} \sum_{I \in \mathcal{I}_n} \sigma^2_{\beta,I} \left(\xi^2_{\beta,I} - 1 \right) \frac{1}{v_{\beta'}} \sum_{I' \in \mathcal{I}_n} \sigma^2_{\beta',I'} \left(\xi^2_{\beta',I'} - 1 \right) = \\ &= \frac{1}{v_\beta} \frac{1}{v_{\beta'}} \sum_{I \in \mathcal{I}_n} \sum_{I' \in \mathcal{I}_n} \sigma^2_{\beta,I} \sigma^2_{\beta',I'} \left[4r^2(\beta, I, \beta', I') - 2r(\beta, I, \beta', I') \right] \end{aligned}$$

and hence by (69) and (iii) of Lemma 5.2 we obtain

$$|\mathbf{E}\,\zeta_\beta\zeta_{\beta'}| \le \frac{Ch^2}{(1-\rho)^2}\frac{1}{v_\beta}\frac{1}{v_{\beta'}}\sum_{I\in\mathcal{I}_n}\sum_{I'\in\mathcal{I}_n}\sigma^2_{\beta,I}\sigma^2_{\beta',I'} \le \frac{Ch}{(1-\rho)^2}$$

and (72) follows.

To prove (74) we note that

$$\begin{aligned} r &= \mathbf{E}\,\xi_{\beta,I}\xi_{\beta',I'} = \\ &= \mathbf{E}\frac{1}{\sigma_{\beta,I}\sigma_{\beta',I'}}\sum_i g_I(X_i^\top\beta)\,g_{I'}(X_i^\top\beta') = \\ &= \frac{n}{\sigma_{\beta,I}\sigma_{\beta',I'}}\int g_I(x^\top\beta)\,g_{I'}(x^\top\beta')\,\pi(x)\,dx. \end{aligned}$$

Introduce new variables y_1 and y_2 with $x^\top\beta = t_I + hy_1,\ x^\top\beta' = t_{I'} + hy_2$. We have

$$|x|^2 = (t_I + hy_1)^2 + \left|\frac{t_{I'} + hy_2 - \rho\,(t_I + hy_1)}{1-\rho}\right|^2, \tag{75}$$

$$r = \frac{nh^{2s+2}}{\sigma_{\beta,I}\sigma_{\beta',I'}(1-\rho)}\int g(y_1)\,g(y_2)\,\pi_1(|x|^2)\,dy_1\,dy_2.$$

Now we use the Taylor expansion for the function $p(y_1,y_2) = \pi_1\left(|x|^2\right)$ with $|x|^2$ due to (75). This function is continuous differentiable and all first derivatives are bounded by $Ch/(1-\rho)$ with some constant C depending only on the function π_1. Using the equality $\int g(t)\,dt = 0$ and (69) we get

$$|r| \le \frac{Cnh^{2s+2}h}{nh^{2s+1}(1-\rho)^2} = \frac{Ch^2}{(1-\rho)^2}.$$

Now everything is prepared to complete the proof of (65). The results of Lemmas 5.2 and 5.3 reduce this assertion to the following one:

$$\frac{1}{N}\sum_{\beta\in S_n}\left[\exp\left\{v_\beta\tilde{\zeta}_\beta - \frac{1}{2}v_\beta^2\right\} - 1\right] \to 0 \tag{76}$$

under the measure $\mathbf{P}_0$. It suffices to check that

$$\frac{1}{N^2}\mathbf{E}_0\left|\sum_{\beta\in S_n}\left(\tilde{Z}_\beta - 1\right)\right|^2 \to 0$$

with

$$\tilde{Z}_\beta = \exp\left\{v_\beta\tilde{\zeta}_\beta - \frac{1}{2}v_\beta^2\right\}.$$

Using normality of $\tilde{\zeta}_\beta$ and (iii) of Lemma 5.2 one derives

$$\mathbf{E}_0 \left[\exp\left\{ v_\beta \tilde{\zeta}_\beta - \frac{1}{2} v_\beta^2 \right\} - 1 \right]^2 = \exp\left\{ v_\beta^2 \right\} \leq n^a.$$

For different $\beta, \beta' \in S_n$ denote $r = \mathbf{E}_0 \tilde{\zeta}_\beta \tilde{\zeta}_{\beta'}$. Then $\tilde{\zeta}_{\beta'}$ can be represented in the form $\tilde{\zeta}_{\beta'} = r\tilde{\zeta}_\beta + (1-r)\zeta'$ with ζ' independent of $\tilde{\zeta}_\beta$. Now

$$\begin{aligned}
\mathbf{E}_0 \tilde{Z}_\beta \tilde{Z}_{\beta'} &= \mathbf{E}_0 \exp\left\{ (v_\beta + r v_{\beta'}) \tilde{\zeta}_\beta - \frac{1}{2} v_\beta^2 \right\} \exp\left\{ (1-r) v_{\beta'} \zeta' - \frac{1}{2} v_{\beta'}^2 \right\} = \\
&= \exp\left\{ \frac{1}{2} (v_\beta + r v_{\beta'})^2 - \frac{1}{2} v_\beta^2 + (1-r)^2 v_{\beta'}^2 - \frac{1}{2} v_{\beta'}^2 \right\} = \\
&= \exp\left\{ r v_\beta v_{\beta'} - r v_{\beta'}^2 + \frac{1}{2} r^2 (v_\beta^2 + v_{\beta'}^2) \right\}.
\end{aligned}$$

The results of Lemma 5.4 and (iv) of Lemma 5.2 allow us to obtain

$$\mathbf{E}_0 \left(\tilde{Z}_\beta - 1 \right) \left(\tilde{Z}_{\beta'} - 1 \right) = \mathbf{E}_0 \tilde{Z}_\beta \tilde{Z}_{\beta'} + 1 \leq C r \ln n.$$

Finally, by (61), Lemma 5.4 and (iii),(iv) of Lemma 5.2 we derive

$$\begin{aligned}
&\frac{1}{N^2} \mathbf{E}_0 \left| \sum_{\beta \in S_n} \left(\tilde{Z}_\beta - 1 \right) \right|^2 = \\
&= \frac{1}{N^2} \sum_{\beta \in S_n} \mathbf{E}_0 \left(\tilde{Z}_\beta - 1 \right)^2 + \frac{1}{N^2} \sum_{\beta \in S_n} \sum_{\beta' \in S_n, \beta' \neq \beta} \mathbf{E}_0 \left(\tilde{Z}_\beta - 1 \right) \left(\tilde{Z}_{\beta'} - 1 \right) \leq \\
&\leq \frac{1}{N^2} \sum_{\beta \in S_n} e^{v_\beta^2} + \frac{1}{N^2} \sum_{\beta \in S_n} \sum_{\beta' \in S_n, \beta' \neq \beta} \frac{C h \ln n}{|\beta - \beta'|^4} \leq \\
&\leq \frac{1}{N^2} n^a + \frac{C h \ln n}{b_n^4} \to 0
\end{aligned}$$

if a is small enough.

References

[1] Carroll, R. J. and Ruppert, D. and Stefanski, L. A. (1995). Nonlinear Measurement Error Models, *Chapman and Hall, New York.*

[2] Carroll, R. J. and Ruppert, D. (1988). Transformation and weighting in regression, *Chapman and Hall, New York.*

[3] Engle, R. F., Granger, W. J., Rice, J. and Weiss, A. (1986). Semiparametric estimates of the relation between weather and electricity sales, *Journal of the American Statistical Association,* **81**, pp. 310-20

[4] Friedman, F. H. and Stuetzle, W. (1981). A projection pursuit regression, *Journal of the American Statistical Association*, **79**, pp. 599-608

[5] Golubev, G. (1992). Asymptotic minimax regression estimation in additive model, *Problems of Information Transmission,* **28**, pp. 3-15 (in russian)

[6] Green, P. and Silverman, B. W. (1994). The penalized likelihood approach, *Chapman and Hall, London.*

[7] Härdle, W. (1990). Applied nonparametric regression, *Econometric Society Monographs No. 19, Cambridge University Press.*

[8] Härdle, W., Klinke, S. and Turlach, T. A. (1995). XploRe-An interactive statistical computing environment, *Springer Verlag.*

[9] Härdle, W. and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits, *Annals of Statistic,* **4**, pp. 1926-47

[10] Härdle, W. and Horowitz, J. (1994). Testing a parametric model against a semiparametric alternative, *Econometric Theory.*

[11] Hall, P. (1989). On projection pursuit regression, *Annals of Statistic,* **17**, pp. 573-8

[12] Horowitz, J. (1993). Semiparametric and nonparametric estimation of quantal response models, in *G. S. Maddala, C. R. Rao and H. D. Vinod (eds), Handbook of Statistics, Elsevier Science Publishers,* pp. 45-72

[13] Huber, P. J. (1985). Projection pursuit, *Annals of Statistic,* **13**, pp. 435-75

[14] Huet, S., Jolivet, E. and Mésseau, A. (1993). La regression non-lineaire: methodes et applications en biologie, *INRA, Paris,* Chapter 1,3.

[15] Ibragimov, I. A. and Khasminski, R.Z. (1977). One problem of statistical estimation in Gaussian white noise, *Soviet Math. Dokl.,* **236**, pp. 1351-4

[16] Ibragimov, I. A. and Khasminski, R.Z. (1981). Statistical Estimation; Asymptotic Theory, *Springer Verlag.*

[17] INGSTER, YU. I. (1982). Minimax nonparametric detection of signals in white Gaussian noise, *Problems of Information Transmission,* **18**, pp. 130-40

[18] INGSTER, YU. I. (1993). Asymptotically minimax hypothesis testing for nonparametric alternatives I, II, III, *Mathematical Methods of Statistics.* **2**, pp. 85-114

[19] MADDALA, G. (1983). Limited-dependent and qualitative variables in econometrics, *Cambridge University Press.*

[20] MCCULLAGH, P. AND NELDER, J. A. (1989). Generalized Linear Models, *Monographs on Statistics and Applied Probability, 2 edn,* **37**, *Chapman and Hall, London.*

[21] MÜLLER, H. G. (1987). Weighted local regression and kernel methods for nonparametric curve fitting, *Journal of the American Statistical Association,* **82**, pp. 231-8

[22] MÜLLER, H. G. (1988). Nonparametric regression analysis of Longitudinal Data. *Lecture Notes in Statistic,* **19**, pp. 817-29

[23] MÜLLER, H. G. (1991). Smooth optimum kernel estimators near endpoints, *Biometrika,* **78**, pp. 521-30

[24] PROENCA, I. AND RITTER, CHR. (1995). Negative bias in the H-H Statistik, *Computational Statistics,*

[25] SEVERINI, T. A. AND STANISWALIS, J. G. (1994). Quasi-likelihood Estimation in Semiparametric Models, *Journal of the American Statistical Association,* **89**, pp.501-11

[26] SPECKMAN, P. (1988). Kernel smoothing in partial linear models, *Journal of the Royal Statistical Society, Series B,* **50**, pp. 413-46

[27] RICE, J. A. (1986). Convergence rates for partially splined models, *Statistics and Probability Letters,* **4**, pp.203-8

[28] WAND, P. AND JONES, M. P. (1994). Introduction to Kernel Smoothing, *Chapman and Hall.*